Subclonal cooperation drives metastasis by modulating local and systemic immune microenvironments

Michalina Janiszewska [1,2,3,15,20], Doris P. Tabassum[1,16,20], Zafira Castaño[3,4],
Simona Cristea[5,6,7], Kimiyo N. Yamamoto[5,6,7], Natalie L. Kingston[1], Katherine C. Murphy[1],
Shaokun Shu[1,2,3], Nicholas W. Harper[1], Carlos Gil Del Alcazar[1,2,3], Maša Alečković [1,2,3],
Muhammad B. Ekram[1,2,3,17], Ofir Cohen[1,8], Minsuk Kwak[9,10], Yuanbo Qin[3,4,18], Tyler Laszewski[4],
Adrienne Luoma[11], Andriy Marusyk [1,2,3,19], Kai W. Wucherpfennig[11], Nikhil Wagle [1,2,8], Rong Fan [9,10],
Franziska Michor [5,6,7,8,12,13], Sandra S. McAllister[3,4,8,14] and Kornelia Polyak [1,2,3,8,12,13,14]★

Most human tumours are heterogeneous, composed of cellular clones with different properties present at variable frequencies. Highly heterogeneous tumours have poor clinical outcomes, yet the underlying mechanism remains poorly understood. Here, we show that minor subclones of breast cancer cells expressing IL11 and FIGF (VEGFD) cooperate to promote metastatic progression and generate polyclonal metastases composed of driver and neutral subclones. Expression profiling of the epithelial and stromal compartments of monoclonal and polyclonal primary and metastatic lesions revealed that this cooperation is indirect, mediated through the local and systemic microenvironments. We identified neutrophils as a leukocyte population stimulated by the IL11-expressing minor subclone and showed that the depletion of neutrophils prevents metastatic outgrowth. Single-cell RNA-seq of CD45+ cell populations from primary tumours, blood and lungs demonstrated that IL11 acts on bone-marrow-derived mesenchymal stromal cells, which induce pro-tumorigenic and pro-metastatic neutrophils. Our results indicate key roles for non-cell-autonomous drivers and minor subclones in metastasis.

Tumours are mixtures of cells with distinct characteristics[1]. High intra-tumour diversity increases the likelihood of disease progression[2] as different subclones respond differently to microenvironmental cues. Treatment of heterogeneous tumours favours the selection of resistant subclones, leading to therapeutic failure. Heterogeneous tumours also display phenotypes different from those of individual clones; thus, intra-tumour heterogeneity has a significant impact on tumour progression and therapeutic resistance.

Metastatic disease is responsible for most cancer-associated mortality; therefore, understanding the drivers of metastatic progression is key for improving clinical outcomes. Cancer genome sequencing studies have identified limited genetic differences between primary and metastatic tumours and demonstrated extensive subclonal heterogeneity in both primary and distant lesions[3,4]. However, the mechanism(s) by which polyclonal primary tumours produce polyclonal metastases remains elusive. Moreover, several recent studies implicated microenvironmental changes as key mediators of metastatic dissemination and outgrowth[5,6], thus highlighting the role of non-cell-autonomous factors in tumour evolution.

Clonal cooperation drives polyclonal metastasis

We have been investigating the effect of subclonal interactions on tumour phenotypes using a human breast cancer cell line (MDA-MB-468)-derived xenograft model of intra-tumour heterogeneity. We previously established that a minor subclone can drive tumour growth through non-cell-autonomous interactions, supporting long-term subclonal heterogeneity[7]. Briefly, we tested 18 subclones, each expressing a secreted protein implicated in metastasis and angiogenesis, and found that polyclonal tumours with all 18 subclones grew the fastest, whereas only interleukin 11 (IL11) and chemokine (C-C motif) ligand 5 were able to drive tumour growth in monoclonal tumours. We also determined that a mixture of two subclones expressing IL11 and FOS-induced growth factor (FIGF;

also known as VEGFD) was largely able to reproduce this pheno-type. Omission of IL11[+] cells from polyclonal tumours decreased tumour growth, suggesting that IL11 and FIGF may cooperate. In addition, both polyclonal tumours and tumours comprised of only IL11 and FIGF subclones were highly metastatic, but the underlying mechanism remained undefined.

To dissect the molecular basis of this metastasis-driving sub-clonal cooperation, we first investigated the clonality of metas-tases of primary MDA-MB-468 tumours comprising IL11[+] and FIGF[+] driver subclones, as well as neutral subclones. Monoclonal or polyclonal mixtures of green fluorescent protein (GFP) and luciferase-expressing parental cells, red fluorescent protein (RFP) and V5-tagged IL11[+] cells, and RFP[+]FIGF[+] cells were implanted into the mammary fat pads of immunodeficient NOG mice. We monitored the primary-tumour growth by weekly caliper mea-surements and macrometastatic lesions by weekly biolumines-cence imaging. Polyclonal tumours initiated from 5% IL11[+] and 5% FIGF[+]RFP[+] cells with 90% GFP[+] parental cells grew faster and were more metastatic than monoclonal and parental tumours (Fig. 1a–c and Supplementary Table 1). Immunohistochemistry-based quantification of human cytokeratin[+] (CK[+]) cells in the lungs revealed an increased number of metastatic lesions in mice with FIGF[+] primary tumours (Fig. 1d,e) despite small primary tumours. However, most of these were micrometastases, which could only be detected as single cells by immunohistochemistry, whereas the lungs of mice with polyclonal primary tumours were filled with macrometastases emitting high bioluminescence signals (Fig. 1b). The increased metastases by polyclonal tumours were not simply due to their faster growth, as this trend was still observed when primary tumours were surgically removed when they reached a diameter of 1 cm (Supplementary Fig. 1a,b). Furthermore, IL11[+] monoclonal tumours grew faster than parental and neutral clones but were not as metastatic as polyclonal tumours. Thus, macro-metastatic outgrowth is a phenotype unique to polyclonal tumours. Immunofluorescence demonstrated that metastases of polyclonal tumours were also mostly polyclonal, composed of V5-tag[+] IL11 and FIGF driver and GFP[+] parental subclones, the latter of which did not produce significant macrometastases on their own except for rare GFP[+] monoclonal metastases (Fig. 1f,g). These results show that polyclonal tumours behave differently from monoclo-nal tumours and that IL11[+] and FIGF[+] subclones driving tumour growth and dissemination, respectively, cooperate to generate fre-quent large polyclonal metastases. This cooperation is likely to be indirect, as close cell contact between driver and neutral subclones was not commonly observed and clonal distribution within meta-static lesions was random (Fig. 1g,h).

Indirect clonal cooperation via the microenvironment

To explore the mechanisms by which driver subclones cooperate to promote metastasis, we generated polyclonal tumours composed of fluorescently labelled driver clones (10% FIGF–mCherry and 10% IL11–GFP cells) and neutral clones (40% Thy1.1 cell surface marker and 40% cyan fluorescent protein(CFP)-expressing cells). To ensure that a sufficient number of cells of each fraction could be sorted from these tumours, 10% driver clones were used in this experiment. As controls, we established tumours with 100% of each individual clone. The mice were killed when the primary tumours reached approximately 2 cm. The primary and metastatic lesions were dissociated, and fluorescently labelled cancer and unlabelled murine cells were purified by fluorescence-activated cell sort-ing (FACS) and analysed by RNA sequencing (RNA-seq; Fig. 2a, Supplementary Fig. 1c and Supplementary Table 2).

We first investigated whether a common transcriptional pro-gram was induced by the driver subclones in polyclonal tumours. We compared the expression profiles of each subclone puri-fied from polyclonal tumours with that of the parental tumours

(Supplementary Table 3). We found 36 commonly upregulated genes in the fractions from tumours with driver clones (Fig. 2b), including *TMEM173*, which encodes STING—an immune mod-ulator[8]—and several extracellular matrix components. These expression changes suggested that the driver subclones may promote metastasis by altering the tumour microenvironment. We therefore analysed the differences in stromal cells (that is, cells lacking human cancer cell markers) from monoclonal and polyclonal primary tumours, and lungs. Given that the parental cells did not generate metastases, we used lung tissue from mice with 100% parental cell tumours as controls in our compari-sons. Principal component analyses (PCA) demonstrated a trend towards clustering according to the tumour of origin, imply-ing that tumour cells have a significant impact on the stromal gene expression (Supplementary Fig. 2a,b and Fig. 2c,d). Using MetaCore[9], we analysed the pathway and process-network enrich-ment within the upregulated genes in the stromal cells of primary tumours capable of seeding metastases (that is, containing a driver subclone) compared with the stroma of parental tumours (Supplementary Tables 3 and 4). Interestingly, nine of the ten top significantly enriched pathways were immune related (Fig. 2e and Supplementary Fig. 2c). In contrast, all ten top pathways enriched in non-cancer cells from the lungs of mice with polyclonal tumours were related to vesicular transport and cell motility (Fig. 2e). Comparisons of the FIGF and IL11 fractions of poly-clonal tumours with 100% FIGF[+] or IL11[+] tumours also showed upregulation of angiogenesis-related genes (Supplementary Table 3). Immunohistochemistry analysis of CD31 blood and LYVE1 lymphatic vessel markers confirmed this upregulation even in primary tumours (Supplementary Fig. 2d).

We also analysed the relative frequencies of each of the four sub-clones within polyclonal tumours by FACS and found that the IL11[+] and FIGF[+] subclones remained as minor fractions (Supplementary Fig. 2e). Thus, their ability to drive tumour growth and metastasis did not require clonal expansion, indicating a non-cell-autonomous mechanism. Therefore, subclonal cooperation in primary tumours promotes metastasis by modulating the microenvironment of the primary and metastatic lesions.

Polyclonal tumours affect leukocytes

To follow up on our finding of immune-related changes in the stroma of metastatic-capable primary tumours and to analyse leukocytes, we repeated the polyclonal and monoclonal xenograft experiment using luciferase-expressing cells to monitor metas-tases. Immunodeficient NOG mice retained innate immune and myeloid cells (albeit not fully functional), and the myeloid-related pathways were highly upregulated in the primary tumours driving metastases. We assessed the qualitative changes in blood leuko-cytes by FACS at three weeks after tumour initiation and at the six-week experimental endpoint when the primary tumours, lungs and bone marrow were also collected (Fig. 3a and Supplementary Fig. 3a). The most significant differences were detected in neu-trophils and monocytes (Fig. 3b and Supplementary Fig. 3b). The relative fraction of neutrophils was significantly higher in the blood, primary tumours and lungs of mice with polyclonal and IL11[+] monoclonal tumours, compared with mice bearing parental cell tumours, whereas the fraction of CD11c[+] cells was significantly lower in the blood and lungs, macrophages showed a slight decrease in primary tumours and lungs, and monocytes were decreased in the blood and primary tumours (Fig. 3b and Supplementary Fig. 3b).

Neutrophil influx has been associated with IL11 signalling and inflammatory bursts[10]; we therefore analysed CD45[+] bone mar-row cells for IL11RA expression. Approximately 20–30% of the CD45[+] cells of the mice bearing IL11[+] or FIGF[+] monoclonal and polyclonal tumours were IL11RA[+] (Fig. 3c), and the fraction was
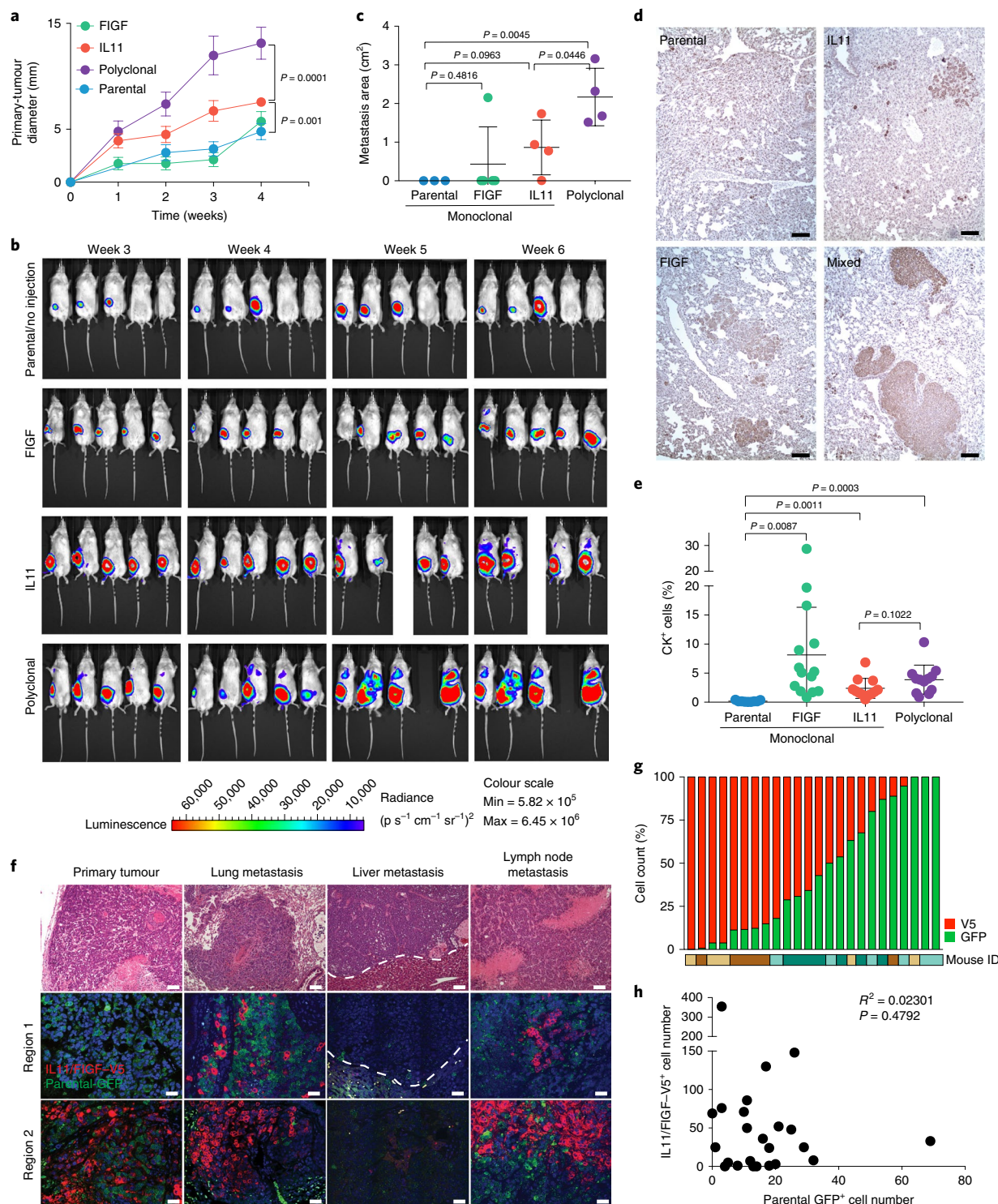
Fig. 1 | Minor driver clones lead to polyclonal metastasis. a, Primary tumour growth kinetics. Data are shown as the mean ± s.d.; $n = 10$ tumours and 5 animals per group. **b**, Representative bioluminescence images. **c**, Quantification of lung macrometastases based on the in vivo bioluminescence at week 6. Data are shown as the mean ± s.d.; $n = 5$ animals per group, except the IL11 and polyclonal groups where $n = 4$ animals. All areas with detected luminescence were quantified and the primary-tumour area was excluded from the analysis. **d**, Representative images of CK staining in the lungs. Scale bars, 100 μm. Staining was repeated twice with similar results. **e**, Quantification of human CK+ cells in the lungs. On average, 3,500 cells per field and three fields per sample were counted. For each group, the lungs of five mice were analysed. The percentage of positive cells per field is shown with the mean ± s.d. **f**, Representative images of the histology (haematoxylin–eosin staining; upper panels) and clonal composition (GFP and V5 immunofluorescence; middle and lower panels) of primary tumours and metastases. Dashed lines demarcate metastatic lesions. Scale bars, 100 μm. Staining was repeated twice with similar results. **g**, Quantification of parental GFP+ and IL11/FIGF-expressing V5+ cells in lung metastatic lesions from four polyclonal tumour-bearing animals, six lesions were quantified per animal. **h**, Frequency of parental GFP+ and IL11/FIGF-expressing V5+ cells per lesion. $n = 24$ lesions in the lungs of 4 animals. The Pearson's coefficient of multiple correlation (R^2) and two-tailed Student's t-test P values are indicated. P values indicate statistical significance based on unpaired two-tailed Student's t-tests for **a,c,e**. See also Supplementary Table 1 for the raw data.
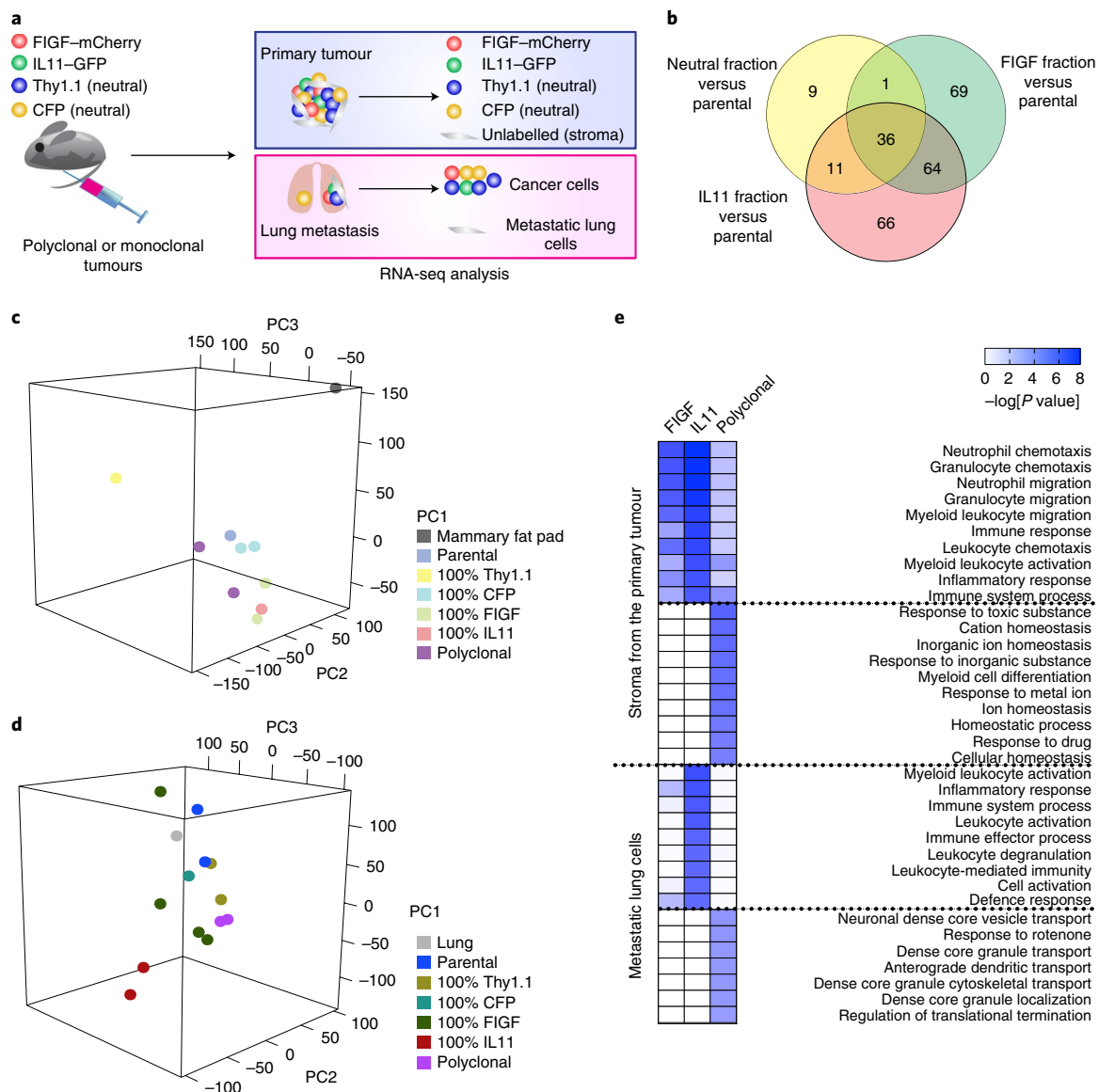
Fig. 2 | Driver-clone-induced changes in the stroma of primary and metastatic tumours. a, Schematic outline of the experimental design. **b,** Overlap between differentially expressed genes in cancer-cell fractions from the polyclonal primary tumours ($n=3$) compared with the parental cell line tumours ($n=3$). **c,d,** Principal component analysis of the RNA-seq profiles of stromal cells from mono- and polyclonal primary tumours (**c**) and non-tumour cells from metastatic lungs (**d**). $n=3$ independent tumours per group. **e,** MetaCore GO Processes overrepresented in the expression profiles of the stroma of primary tumours ($n=3$) and metastasis-bearing lung cells ($n=3$). The colour scale corresponds to $-\log P$ value of the significance of enrichment, which was calculated using the MetaCore enrichment analysis test.

significantly higher in mice with IL11[+] or FIGF[+] monoclonal and polyclonal tumours compared with mice with parental tumours (Fig. 3d). Hence, the presence of minor metastasis-promoting subclones within tumours alters the immune system, particularly the neutrophil population.

In breast cancer patients, distant metastases are usually recurrences after the surgical removal of primary tumours and some studies have suggested that the surgery itself may induce metastatic lesions[11,12]. Thus, we performed mock or primary-tumour-removal surgery in another experiment but did not observe any effects of surgery on the metastatic burden in the lungs (Supplementary Fig. 4a–c). Therefore, surgery-related wound healing and inflammation are not inherently responsible for the observed immune-related changes or increased metastasis in our polyclonal tumours.

IL11–FIGF cooperation drives pro-metastatic systemic changes

Next, we investigated whether the metastasis-promoting subclones exert their effects via local or systemic changes. We injected a 1:1 mixture of MDA-MB-468 cells expressing doxycycline (DOX)-inducible, haemagglutinin (HA)-tagged IL11 and FIGF into the right mammary fat pad and a slow-growing, non-metastatic patient-derived xenograft (PE1) into the contralateral mammary fat pad. The mice were randomized into −/+ DOX groups to induce IL11 and FIGF expression. As expected, the IL11[+]FIGF[+] tumours grew faster than parental tumours, although the difference was not as profound as in polyclonal tumours with constitutive expression. The growth of the PE1 primary tumours was not affected by the contralateral IL11[+]FIGF[+] tumour (Fig. 4a,b), but lung metastases were significantly increased in the +DOX mice (Fig. 4c,d). Using
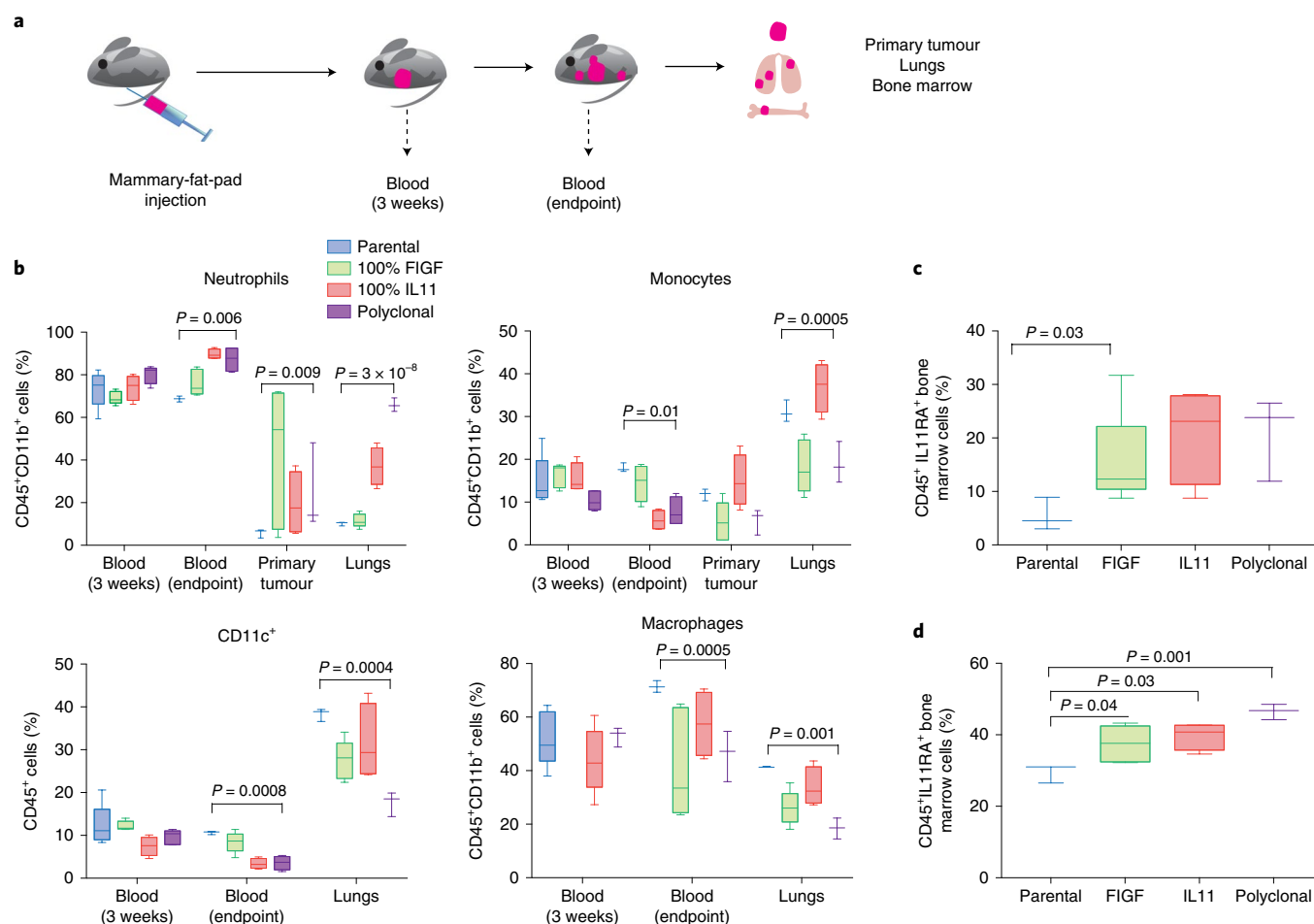
Fig. 3 | The effect of polyclonal tumours on the immune system. a, Schematic outline of the experimental design. **b**, Analysis of the myeloid cells from the blood, primary tumours and lungs of mice bearing parental, 100% FIGF and 100% IL11 monoclonal, and polyclonal tumours; $n = 5$ per group. Legend applies to all graphs. **c**, Fraction of CD45$^+$ bone marrow cells expressing IL11RA; $n = 5$ per group. **d**, Fraction of neutrophils among the CD45$^+$IL11RA$^+$ bone marrow cells; $n = 5$ per group. See also Supplementary Table 1 for the raw data. The box-and-whisker plots in all bar graphs show the mean (midline), and 25th–75th (box) and 5th–95th (whiskers) percentiles. P values were calculated using an unpaired two-tailed Student's t-test.

immunohistochemistry for the HA tag, we found that the metastatic lesions were a mixture of HA$^+$ and HA$^-$ cells, with the latter presumably representing the PE1 cells (Fig. 4c). To confirm that these negative cells were not MDA-MB-468 cells that had lost the expression of HA-tagged IL11 and FIGF, we performed fluorescence in situ hybridization (FISH) to the epidermal growth factor receptor (*EGFR*) and the human-specific CEP8, as MDA-MB-468 cells have an *EGFR* amplification that is not detected in PE1, whereas CEP8 marks all human cells. This analysis clearly demonstrated that metastases were a mixture of PE1 and *EGFR*^{amp} IL11$^+$/FIGF$^+$ MDA-MB-468 cells (Fig. 4e). Single-cell quantification of the FISH signal showed a significant correlation between the numbers of IL11$^+$/FIGF$^+$ (*EGFR*^{amp}) and PE1 (CEP8^{only}) cells, implying potential local interactions between these populations (Fig. 4f). Therefore, IL11 and FIGF promote metastasis via systemic effects, potentially mediated by neutrophils.

Depletion of neutrophils affects metastases
We treated mice with a neutrophil-depleting Ly6G antibody[13] to determine whether neutrophils are required for the increased metastasis driven by IL11 and FIGF. To determine the optimal timing to initiate treatment, we used mathematical modelling to predict the onset of metastasis based on our bioluminescence data (see Methods for details). Our computational framework considers two

types of cells—those residing in the primary tumour and those that have left the primary site (Supplementary Fig. 4d). Each time a cell leaves (at a metastatic rate q), it establishes a new metastatic colony. Growth and dissemination rates were estimated from experimental data (Supplementary Fig. 4b) dependent on the experimental condition (Supplementary Fig. 4e). Briefly, to estimate the metastatic rate q, we minimized the distance between the experimental results and model predictions for the distribution of the size of the first metastasis six weeks after the start of the experiment and the primary-tumour size when the first metastasis was detected. We identified the metastatic rate as $10^{-7} < q < 10^{-6}$ per cell per week and, using this, defined the timing of the first metastatic cell leaving the primary site as 5.18 ± 2.45 d for IL11$^+$ monoclonal tumours and 5.11 ± 2.45 d for polyclonal tumours when $q = 10^{-7}$, and 1.33 ± 1.05 and 1.68 ± 0.91 d, respectively, when $q = 10^{-6}$ (Supplementary Fig. 4f–k). Thus, we concluded that treatment should be initiated immediately after tumour inoculation to prevent metastatic seeding.

We initiated polyclonal xenografts with 10% DOX-inducible IL11$^+$ and FIGF$^+$, and 90% parental cells and assigned animals to receive DOX with a neutrophil-blocking anti-Ly6G antibody or an isotype control, which was administered for four weeks (Supplementary Fig. 5a). We confirmed the efficacy of the anti-Ly6G antibody treatment after two weeks by assessing the blood neutrophil counts (Supplementary Fig. 5b–d). At six weeks, we
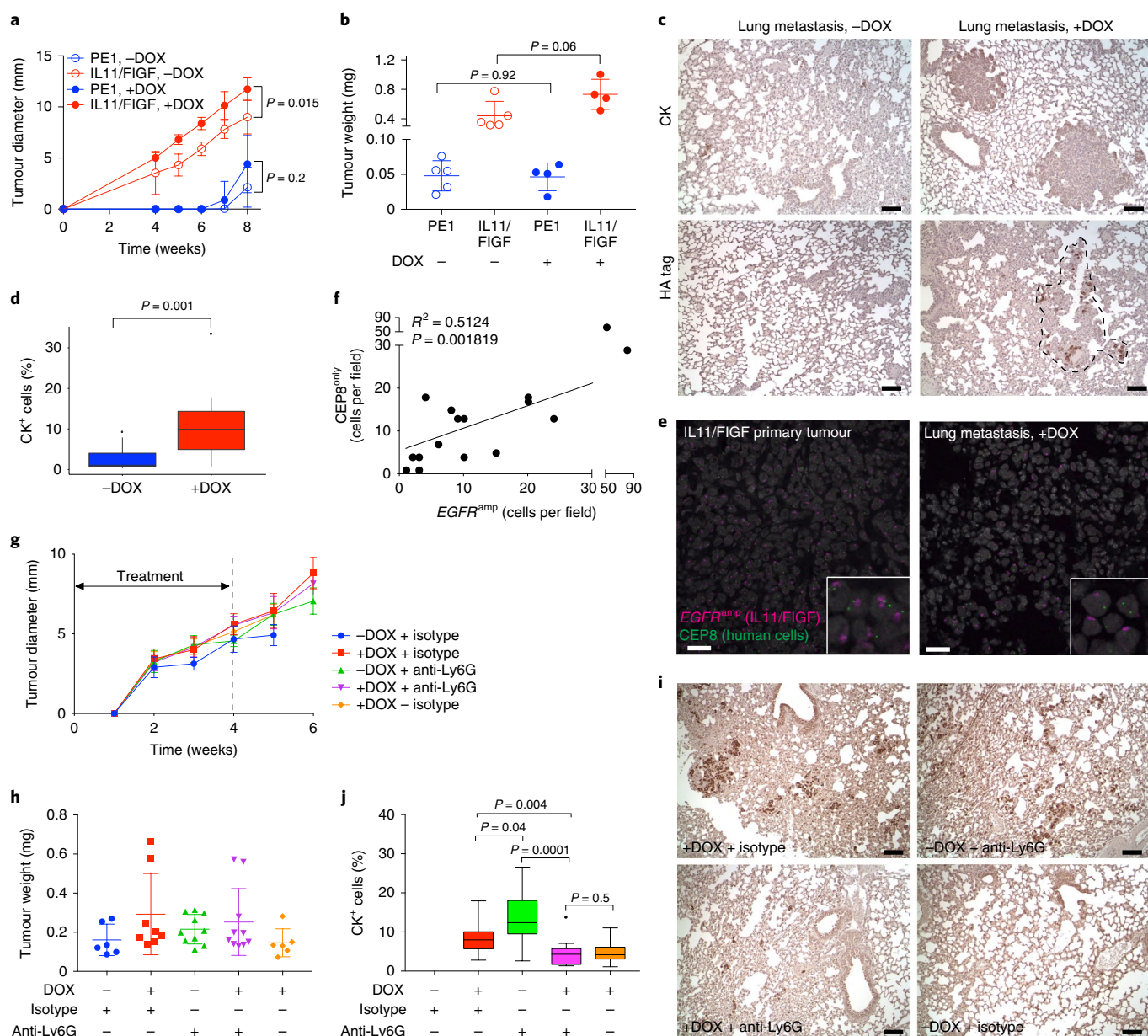
Fig. 4 | Systemic effects of metastasis-driver subclones. a–d, Effects of driver subclones (IL11 and FIGF under a DOX-inducible promoter) on the growth and metastasis of contralaterally injected patient-derived xenografts. **a,b,** Primary tumour growth kinetics (**a**) and final tumour weight (**b**). Data are shown as the mean ± s.d. of two independent experiments; $n = 5$ animals per group. Two-tailed unpaired Student's t-test P values are shown. **c,** Representative immunohistochemical images of the indicated staining. Scale bars, 100 μm. Dashed line demarcates a metastatic lesion. **d,** Fraction of CK$^+$ cells. On average, 2,900 cells per field and three fields per sample were counted. For each group, the lungs of five mice were analysed. Data shown are representative of two independent experiments. The box-and-whisker plots show the mean (midline), 25th–75th (box) and 5th–95th (whiskers) percentiles, and outliers (dots). The two-tailed unpaired Student's t-test P value is shown. **e,** Representative images of EGFR and CEP8 FISH on primary tumour and lung metastasis under DOX treatment. Scale bars, 25 μm. Staining was repeated twice with similar results. **f,** Correlation between the numbers of cells per metastatic lesion that are EGFR^{amp} or have CEP8^{only} FISH signal. Three or four fields of view of the lungs from four animals were quantified; a total of 16 lesions were analysed. The Pearson's correlation R^2 and two-tailed Student's t-test P values are indicated. **g–j,** Neutrophil depletion in mice bearing polyclonal tumours with minor driver subclones. **g,h,** Tumour growth kinetics (**g**) and final tumour weight (**h;** $n = 5$ per group). The dashed line in **g** marks the anti-Ly6G and isotype-antibody treatment duration. Data are shown as the mean ± s.d. **i,** Representative images of CK staining of the lungs of the treated mice. Scale bars, 100 μm. **j,** Quantification of CK$^+$ cells. On average, 2,600 cells per field and three fields per sample were counted. For each group, the lungs of five mice were analysed. The box-and-whisker plots show the mean (midline), 25th–75th (box), and 5th–95th (whiskers) percentiles, and outliers (dots). Two-tailed unpaired Student's t-test P values are shown. See also Supplementary Table 1 for the raw data.

found that neutrophil depletion did not affect the growth of the primary tumour (Fig. 4g,h) but significantly decreased the proportion of CK$^+$ metastatic cells in the lungs of mice compared with mice treated with the isotype control (Fig. 4i,j). The comparison to untreated animals achieved borderline significance. Unexpectedly, anti-Ly6G treatment increased the metastatic load in −DOX mice with tumours that did not express IL11 and FIGF (Fig. 4i,j). These results demonstrate that the systemic depletion of neutrophils

decreases metastatic spread (or outgrowths) of polyclonal tumours driven by IL11+ and FIGF+ subclones, and can have opposing effects on metastases depending on factors produced by the primary tumour. Furthermore, the injection of an isotype-control antibody can enhance metastasis. Therefore, the phenotype and functional properties of neutrophils depend on the systemic host environment.

Single-cell analysis of CD45+ cell populations

To further investigate the effects of the IL11+ and FIGF+ subclones on leukocytes, we repeated the above experiment with polyclonal xenografts composed of 10% DOX-inducible IL11+ and FIGF+, and 90% parental cells. We purified CD45+ cells from the blood, primary tumours and lungs of the mice by FACS six weeks after tumour inoculation, and performed single-cell RNA-seq (scRNA-seq)[13]. Clustering analysis revealed a minor fraction of contaminating CK+EGFR+ cancer cells, that were removed from subsequent analyses (Supplementary Fig. 6a–c), as well as numerous identifiable leukocyte subpopulations (Fig. 5a, Supplementary Fig. 6d–g and Supplementary Table 5). CD45+ cells from the lungs, blood and primary tumours differed in their contribution to the distinct cell clusters (Fig. 5b). For example, M1 macrophages were found only in primary tumours, whereas M2 macrophages were present only in lung metastases. Given that our previous results indicated that neutrophils contribute to metastasis, we compared the single-cell transcriptomic profiles of blood and lung neutrophils from mice with or without DOX treatment. The IL11- and FIGF-induced upregulation of several pathways—including TGFβ and JAK–STAT signalling—associated with pro-tumorigenic and pro-metastatic immune signatures in neutrophils from the lungs but not from blood (Fig. 5c and Supplementary Tables 6,7). Despite the observed change in lung neutrophils, we did not detect IL11- or FIGF-receptor expression in this cell population through scRNA-seq. However, cells with *IL11RA* transcripts were clearly present in a separate cluster (Fig. 5d), which could not be classified as either neutrophils or any other leukocyte subpopulation. These *IL11RA+* cells express *IL6ST*, which encodes GP130, a coreceptor necessary for IL11 signalling, and STAT3, a downstream IL11 effector (Supplementary Fig. 6f). Based on gene expression in this cluster (Supplementary Table 8), which also included extracellular-matrix and development-related proteins, we denoted this population as IL11-responsive mesenchymal stromal cells (MStrCs). Although this population does not express the classic mesenchymal stem cell (MSC) markers, its signature is rich in stem-cell-related genes, implying that it could be a previously uncharacterized MSC-like cell type. Previous studies have described interactions between MSCs and leukocytes mediated by various cytokines and chemokines[14]. Thus, we identified secreted factors from our MStrC-specific expression signatures using gene-enrichment analysis and analysed them in different conditions (Fig. 5e). Comparison of lung MStrCs from −/+ DOX animals revealed that the induction of IL11/FIGF expression in the primary tumours upregulated the expression of chemoattractants for pro-metastatic neutrophils including *CXCL12*, *CXCL14* and *CXCL1* (Fig. 5e).

We performed immunofluorescence and FACS to validate the presence of CD45+IL11RA+ cells in the lungs of mice and to assess co-expression of MStrC markers with IL11RA . CD45+IL11RA+ cells were relatively common in the spleen and present as a rare subpopulation with mesenchymal morphology in the lungs (Fig. 5f). Rare IL11RA[high] cells co-expressing PLXDC2 and ANTRXR1—markers of these cells based on scRNA-seq (Supplementary Table 8)—representing 0.5–1% of cells were also identified in the lungs of mice by FACS (Fig. 5g and Supplementary Fig. 6h–l). Immunofluorescence analysis of the mice lungs demonstrated the presence of rare IL11RA+PLXDC2+ cells with mesenchymal morphology (Fig. 5h). Lastly, we sorted cells from the lungs of mice with IL11+/FIGF+ tumours by IL11RA expression and analysed their gene expression profiles. Our MStrC signature was significantly enriched in

IL11RA[high] samples (Fig. 5i), and enrichment for cell adhesion and extracellular-matrix remodelling-related process networks was noted (Supplementary Table 9).

In summary, our scRNA-seq data and its subsequent validation by FACS, immunofluorescence and bulk RNA-seq, uncovered an interaction loop among different subpopulations of bone-marrow-derived cells that mediates the pro-metastatic effects of tumour cell-secreted IL11 and FIGF (Fig. 5j). The expression of IL11 and FIGF in a minor subpopulation of cancer cells in polyclonal tumours alters these MStrCs in the lungs, leading to the activation of 'effector' neutrophils that are crucial for metastatic progression.

IL11-driven signatures in human metastatic breast cancer

To investigate the clinical relevance of our findings, we analysed the expression of IL11 and IL11-driven neutrophil and MStrC signatures in breast cancer metastases. IL11 is amplified in a subset of primary breast tumours and this amplification is more commonly observed in metastases (cBioPortal[15,16]; Fig. 6a). Based on our data, IL11 and FIGF can drive significant systemic changes even when expressed by minor subclones, which makes them difficult to detect. We thus explored whether gene signatures of IL11/FIGF-activated neutrophils and MStrCs, defined by our single-cell RNA data, are present in distant metastases in the Metastatic Breast Cancer Project (MBCP) dataset[17], which contains RNA expression data of 156 distant metastatic lesions. As most of the systemic changes were IL11-driven in our experimental system, we classified metastatic lesions into IL11[high] and IL11[low] groups based on the messenger RNA levels. We then compared the IL11[high] and IL11[low] metastases for the expression of gene signatures activated by IL11/FIGF in the lung and blood neutrophils and in MStrCs (Supplementary Table 6). The signatures for the activated lung and blood neutrophils, and lung MStrCs were significantly higher in IL11[high] than in IL11[low] metastases (Fig. 6b). In contrast, the gene signature of IL11/FIGF-activated cancer cells, derived from genes expressed by neutral clones in polyclonal tumours, did not correlate with IL11 expression in metastases (Fig. 6b, bottom panel). Analysis of the expression of these gene signatures in matched primary and metastatic lesions using the MBCP dataset showed that activated neutrophil and MStrC signatures were significantly higher in metastases, whereas cancer cell signatures did not show a difference (Fig. 6c, left). To eliminate potential bias due to differences in tissue type used for RNA extraction (formalin-fixed paraffin-embedded for primary tumours and frozen for metastases), we also compared the MBCP metastases to ER+ tumours in the The Cancer Genome Atlas (TCGA) dataset with essentially the same findings (Fig. 6c, right). Therefore, IL11/FIGF-driven microenvironmental changes are also detected in human breast cancer patients and are associated with distant metastases.

Discussion

Polyclonal metastasis of breast cancer has been described in several recent studies[18,19]. Here we provide evidence that minor subpopulations in polyclonal tumours can drive metastatic progression and outgrowth of polyclonal metastases that also contain neutral subclones (non-metastatic on their own) by modulating the local and systemic microenvironments.

In our study, we focused on two secreted factors, selected based on our previous data[7], which are expressed by minor subclones that cooperate to drive metastasis. IL11 belongs to the IL6 family of cytokines and has been shown to play a role in progression and resistance of multiple cancer types, including prostate and colon cancer[20,21]. In breast cancer, IL11 has been implicated in therapeutic resistance[22] and bone metastasis[23], as well as a marker of poor outcomes[24]. FIGF is a ligand for VEGFR2 and VEGFR3, and can stimulate both angiogenesis and lymphangiogenesis[25,26]. Our results show that FIGF enhances dissemination, as evidenced by the increased numbers of CK+ cells in the lungs, but is not sufficient to drive
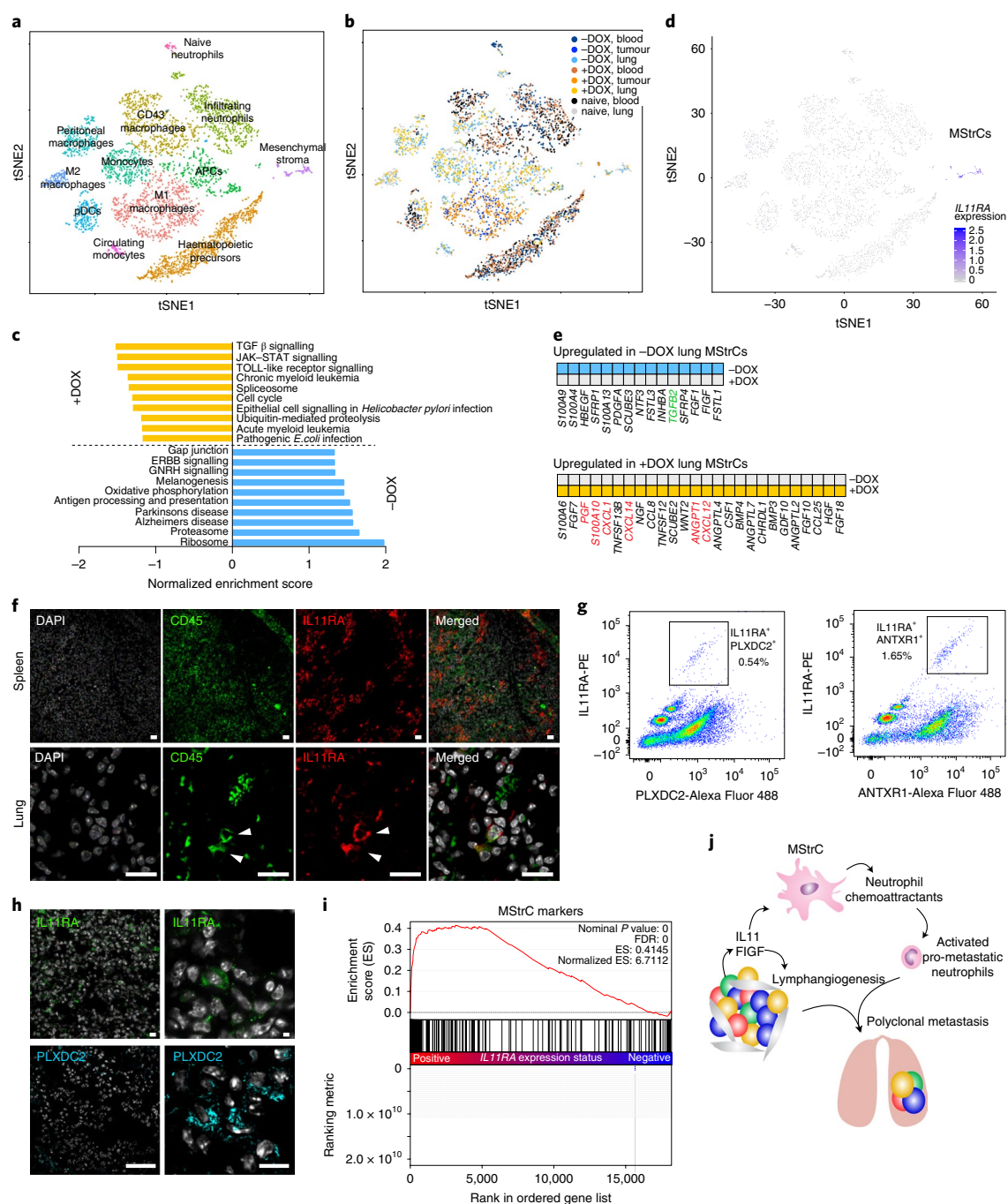
Fig. 5 | Analysis of scRNA-seq data of CD45+ cells. a,b, T-distributed stochastic neighbour embedding (tSNE) clustering of single-cell expression profiles. Each dot represents a single cell ($n = 7,704$) and the distance between cells is a measure of their transcriptional similarity. **a,** Twelve distinct clusters of immune cells were identified according to expression of cell type-specific marker genes or expression signatures (Supplementary Fig. 6 and Supplementary Table 5). The colours denote the distinct clusters of cells. APC, antigen-presenting cells. **b,** Contribution of cells from different samples to the distinct clusters. The cells are coloured according to the sample of origin. **c,** Gene-expression signature analysis of neutrophils from the lungs of mice with (+DOX) or without (−DOX) IL11/FIGF-expressing tumours. **d,** The expression of *IL11RA* was analysed in individual cells ($n = 7,704$). **e,** Secreted factors among cluster-specific differentially expressed genes in lung MStrCs. The genes highlighted in green and red are characteristic of anti-tumorigenic neutrophils and neutrophil chemoattractants, respectively. **f,** Immunofluorescence analysis of CD45 and IL11RA expression in the spleens and lungs of mice with IL11/FIGF-expressing tumours. The white arrows highlight rare CD45+IL11RA+ cells with mesenchymal morphology in the lungs. Staining was repeated twice with similar results. Scale bars, 25 μm. **g,** Flow cytometry analysis of the lungs of tumour-free mice stained with AF55-conjugated IL11RA, and AF488-conjugated PLXDC2 and ANTXR1 (TEM8) antibodies. The highlighted populations show IL11RA^hiPLXDC2+ as well as IL11RA^hiANTXR1+ double-positive cells. **h,** Immunofluorescence analysis depicting rare IL11RA+ and PLXDC2+ cells with mesenchymal morphology in the lungs of +DOX-isotype control-treated mice. Staining was repeated twice with similar results. Scale bars, 25 μm. **i,** Mesenchymal stromal cell signature enrichment in cells from the lungs of tumour-bearing mice, which were FACS sorted for IL11RA expression from $n = 3$ animals. P values of the gene set enrichment analysis (Wald test with Benjamini–Hochberg's multiple-comparison correction) are shown. FDR, false discovery rate. **j,** Model of metastatic cascade driven by IL11/FIGF-expressing minor primary tumour subclones.
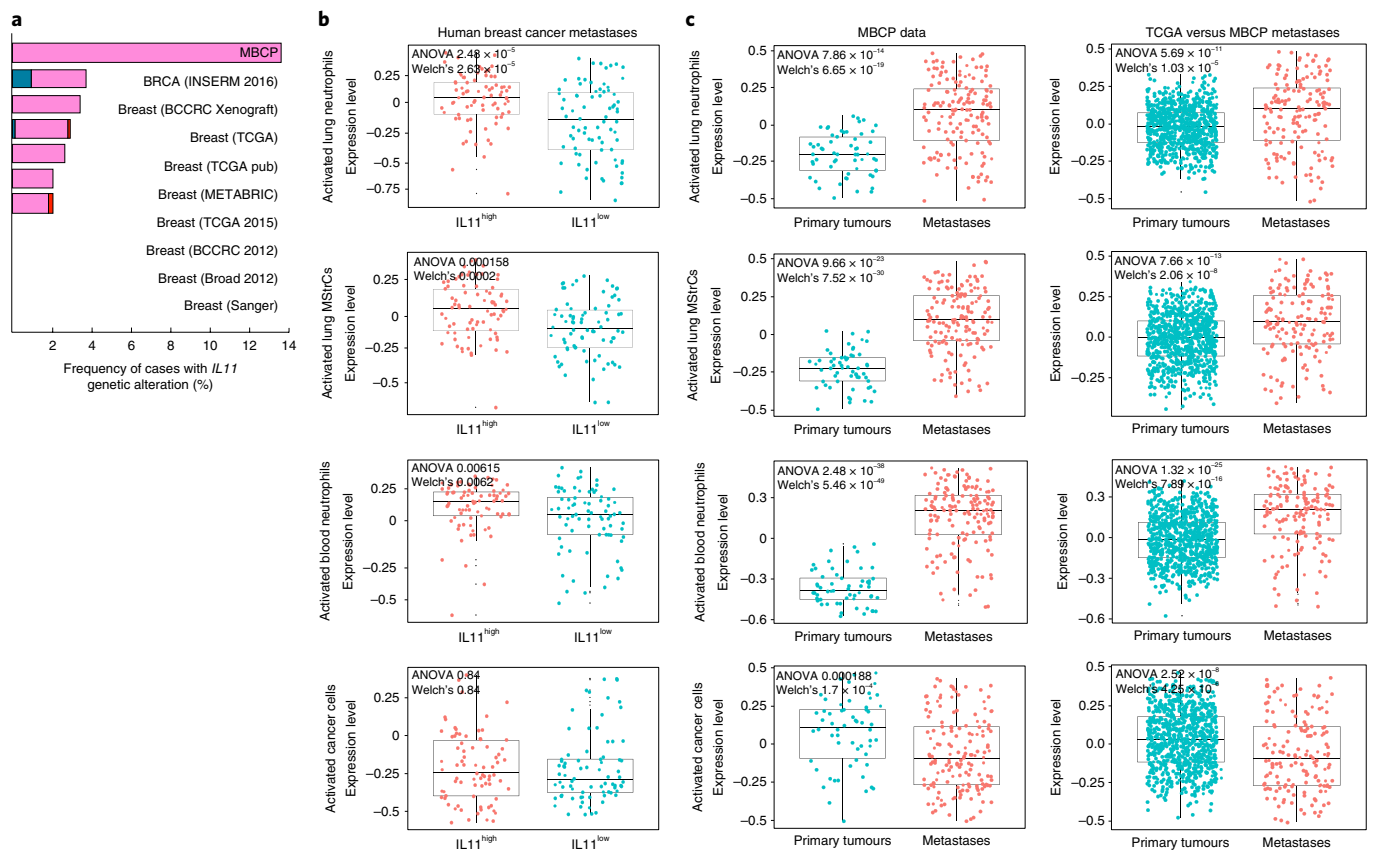
Fig. 6 | IL11 is associated with metastasis in breast cancer patients. a, Frequency of genetic alterations of the *IL11* locus in the cBioPortal breast cancer dataset. **b**, Expression of IL11/FIGF-activated signatures of lung and blood neutrophils, MStrCs and primary cancer cells in human breast cancer metastases with high and low IL11 levels. *n* = 156. The cancer cell signature was derived from a neutral clone isolated from polyclonal primary tumours. **c**, Gene expression signatures of IL11/FIGF-activated cells in the MBCP dataset (matched primary (*n* = 58) and metastatic lesions (*n* = 156); left) and in the ER$^+$ TCGA primary tumours (*n* = 868) versus MBCP metastatic lesions (*n* = 156; right). The box-and-whisker plots show the mean (midline), and 25th–75th (box) and 5th–95th (whiskers) percentiles. *P* values determined using analysis of variance (ANOVA) and two-sided Welch's tests are shown. See also Supplementary Table 1 for the raw data.

macrometastatic outgrowths. In our model, IL11 and FIGF secreted by minor subclones have profound effects on the metastatic ability of breast cancer cells. Although IL11 acts on leukocytes to induce metastasis-permissive changes in the microenvironment of distant organs, FIGF acts more locally by increasing lymphangiogenesis and vascular permeability. However, because these subclones represent only 1–5% of the tumour cells, their presence and their secreted factors may be difficult to detect in clinical samples, making it challenging to predict their metastasis-promoting effects in human tumours. Nevertheless, our analysis of human metastatic breast tumours suggests that the systemic changes in immune and other bone-marrow-derived cells, which are induced by even low levels of these secreted factors, can be detected and potentially exploited for patient stratification and treatment.

We identified neutrophils as the leukocyte population required for IL11- and FIGF-induced polyclonal metastasis. A caveat of our approach is the use of immunocompromised animals, which did not allow us to study the roles of lymphoid cells in metastatic cooperation. Thus, it is possible that other immune cells may also play a role in the metastatic process. Neutrophils have been implicated in metastatic progression by promoting the pre-metastatic niche[27–31]. Furthermore, a high pre-operative neutrophil-to-lymphocyte ratio is associated with poor prognosis in breast cancer patients[32]. Neutrophils have also been shown to have anti-tumour effects and to inhibit seeding of cancer cells in the lungs[33,34]. In our experiments,

depending on whether or not the tumours expressed IL11 and FIGF, the depletion of neutrophils could either decrease or enhance lung metastasis. These seemingly contradictory roles for neutrophils could potentially be explained by our observation, based on scRNA-seq data, that there are several different types of neutrophils, potentially pro- and anti-tumorigenic, depending on the anatomical location and gene expression patterns. Our findings are therefore in agreement with reports that the context and temporal infiltration of monocytes and neutrophils is critical to outcome[31,35,36], and that heterogeneity of not only tumour cells but also immune cells must be considered. Another point to consider is that the injection of isotype-control antibody unexpectedly increased the metastatic burden, potentially by triggering an inflammatory response. The magnitude of the neutrophil-depletion effect on metastasis may therefore be very sensitive to the systemic environment of the host.

The use of single-cell expression profiling of CD45$^+$ cells in our study allowed us to uncover interactions between different subpopulations. We found that leukocytes may not be the direct cellular targets of IL11, but may be indirectly affected through factors secreted by MStrCs responsive to IL11. Interestingly, these cells also express *PLXDC2* and *ANTXR1*, which are highly expressed in tumour-associated endothelial cells[37,38]. Thus, these IL11RA$^+$ mesenchymal cells may represent progenitors that can give rise to multiple cell types. Further characterization of these cells and their effector leukocytes will shed new light on the biology of pro-tumorigenic neutrophils.

An improved understanding of neutrophil subtypes and the development of tools to target pro-metastatic neutrophils, in combination with cancer cell-targeting drugs, could potentially be used to prevent breast cancer metastasis.

References

1. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* **12**, 323–334 (2012).
2. Burrell, R. A. & Swanton, C. Re-evaluating clonal dominance in cancer evolution. *Trends Cancer* **2**, 263–276 (2016).
3. Yates, L. R. et al. Genomic evolution of breast cancer metastasis and relapse. *Cancer Cell* **32**, 169–184 (2017).
4. Brastianos, P. K. et al. Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5**, 1164–1177 (2015).
5. Peinado, H. et al. Pre-metastatic niches: organ-specific homes for metastases. *Nat. Rev. Cancer* **17**, 302–317 (2017).
6. McAllister, S. S. & Weinberg, R. A. The tumour-induced systemic environment as a critical regulator of cancer progression and metastasis. *Nat. Cell Biol.* **16**, 717–727 (2014).
7. Marusyk, A. et al. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* **514**, 54–58 (2014).
8. Chen, Q., Sun, L. & Chen, Z. J. Regulation and function of the cGAS–STING pathway of cytosolic DNA sensing. *Nat Immunol.* **17**, 1142–1149 (2016).
9. Nikolsky, Y., Nikolskaya, T. & Bugrim, A. Biological networks and analysis of experimental data in drug discovery. *Drug Discov. Today* **10**, 653–662 (2005).
10. Ernst, M. & Putoczki, T. L. Molecular pathways: IL11 as a tumor-promoting cytokine-translational implications for cancers. *Clin. Cancer Res.* **20**, 5579–5588 (2014).
11. Retsky, M., Demicheli, R., Hrushesky, W., Baum, M. & Gukas, I. Surgery triggers outgrowth of latent distant disease in breast cancer: an inconvenient truth? *Cancers* **2**, 305–337 (2010).
12. Tohme, S., Simmons, R. L. & Tsung, A. Surgery for cancer: a trigger for metastases. *Cancer Res.* **77**, 1548–1552 (2017).
13. Daley, J. M., Thomay, A. A., Connolly, M. D., Reichner, J. S. & Albina, J. E. Use of Ly6G-specific monoclonal antibody to deplete neutrophils in mice. *J. Leukoc. Biol.* **83**, 64–70 (2008).
14. Turley, S. J., Cremasco, V. & Astarita, J. L. Immunological hallmarks of stromal cells in the tumour microenvironment. *Nat. Rev. Immunol.* **15**, 669–682 (2015).
15. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
16. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
17. Jeselsohn, R. et al. Allele-specific chromatin recruitment and therapeutic vulnerabilities of ESR1 activating mutations. *Cancer Cell* **33**, 173–186 (2018).
18. Siegel, M. B. et al. Integrated RNA and DNA sequencing reveals early drivers of metastatic breast cancer. *J. Clin. Invest.* **128**, 1371–1383 (2018).
19. Hoadley, K. A. et al. Tumor evolution in two patients with basal-like breast cancer: a retrospective genomics study of multiple metastases. *PLoS Med.* **13**, e1002174 (2016).
20. Onnis, B., Fer, N., Rapisarda, A., Perez, V. S. & Melillo, G. Autocrine production of IL-11 mediates tumorigenicity in hypoxic cancer cells. *J. Clin. Invest.* **123**, 1615–1629 (2013).
21. Putoczki, T. L. et al. Interleukin-11 is the dominant IL-6 family cytokine during gastrointestinal tumorigenesis and can be targeted therapeutically. *Cancer Cell* **24**, 257–271 (2013).
22. Bockhorn, J. et al. MicroRNA-30c inhibits human breast tumour chemotherapy resistance by regulating TWF1 and IL-11. *Nat. Commun.* **4**, 1393 (2013).
23. Kang, Y. et al. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**, 537–549 (2003).
24. Hanavadi, S., Martin, T. A., Watkins, G., Mansel, R. E. & Jiang, W. G. Expression of interleukin 11 and its receptor and their prognostic value in human breast cancer. *Ann. Surg. Oncol.* **13**, 802–808 (2006).
25. Bower, N. I. et al. Vegfd modulates both angiogenesis and lymphangiogenesis during zebrafish embryonic development. *Development* **144**, 507–518 (2017).
26. Van den Eynden, G. G. et al. Comparison of molecular determinants of angiogenesis and lymphangiogenesis in lymph node metastases and in primary tumours of patients with breast cancer. *J. Pathol.* **213**, 56–64 (2007).
27. Leach, J., Morton, J. P. & Sansom, O. J. Neutrophils: homing in on the myeloid mechanisms of metastasis. *Mol. Immunol.* **110**, 69–76 (2017).
28. Wculek, S. K. & Malanchi, I. Neutrophils support lung colonization of metastasis-initiating breast cancer cells. *Nature* **528**, 413–417 (2015).
29. Coffelt, S. B. et al. IL-17-producing γδ T cells and neutrophils conspire to promote breast cancer metastasis. *Nature* **522**, 345–348 (2015).
30. Fridlender, Z. G. et al. Polarization of tumor-associated neutrophil phenotype by TGF-beta: "N1" versus "N2" TAN. *Cancer Cell* **16**, 183–194 (2009).
31. Castano, Z. et al. IL-1β inflammatory response driven by primary breast cancer prevents metastasis-initiating cell colonization. *Nat. Cell Biol.* **20**, 1084–1097 (2018).
32. Krenn-Pilko, S. et al. The elevated preoperative derived neutrophil-to-lymphocyte ratio predicts poor clinical outcome in breast cancer patients. *Tumour Biol.* **37**, 361–368 (2016).
33. Granot, Z. et al. Tumor entrained neutrophils inhibit seeding in the premetastatic lung. *Cancer Cell* **20**, 300–314 (2011).
34. Finisguerra, V. et al. MET is required for the recruitment of anti-tumoural neutrophils. *Nature* **522**, 349–353 (2015).
35. Voloshin, T. et al. Blocking IL1β pathway following paclitaxel chemotherapy slightly inhibits primary tumor growth but promotes spontaneous metastasis. *Mol. Cancer Ther.* **14**, 1385–1394 (2015).
36. Kersten, K. et al. Mammary tumor-derived CCL2 enhances pro-metastatic systemic inflammation through upregulation of IL1β in tumor-associated macrophages. *Oncoimmunology* **6**, e1334744 (2017).
37. St Croix, B. et al. Genes expressed in human tumor endothelium. *Science* **289**, 1197–1202 (2000).
38. Carson-Walter, E. B. et al. Cell surface tumor endothelial markers are conserved in mice and humans. *Cancer Res.* **61**, 6649–6655 (2001).

Author contributions

D.P.T. and M.J. performed the xenograft, molecular profiling and immunohistochemical experiments, and data analyses. M.B.E. and N.W.H. analysed the RNA-seq data. A.M. helped with the study conception and xenograft experiments. N.L.K. and K.C.M. assisted with the immunohistochemical staining. M.K. generated the low-input RNA-seq libraries. Y.Q., Z.C., M.A. and C.G.D.A. performed the FACS analyses. T.L. and S.S. assisted with the animal experiments. K.N.Y. carried out the mathematical modelling. A.L. and K.W.W. assisted with the generation of the scRNA-seq libraries. S.C. analysed the scRNA-seq data. O.C. and N.W. provided the MBCP cohort data and performed analyses. K.P. supervised the research with help from F.M., S.S.M. and R.F. All authors helped to design the study and write the manuscript.

Competing interests

The authors declare competing financial interests. K.P. received research support from and was a consultant to Novartis Oncology during the execution of this study. K.P. also serves on the Scientific Advisory Board of Mitra Biotech.

Additional information

Methods

Cell lines and tissue culture conditions. The parental MDA-MB-468 cell line was used to generate IL11- and FIGF-overexpressing subpopulations using lentiviral vectors as previously described[7]. Inducible IL11 and FIGF expression was achieved by cloning the respective complementary DNA into pLentiX2 vectors, followed by viral transduction of the respective constructs into parental MDA-MB-468 cells[7] and selection with $2\,\mu g\,ml^{-1}$ puromycin. All cell lines were cultured in McCoy's medium (Cell Grow) with 10% fetal bovine serum (FBS; HyClone) and $10\,\mu g\,ml^{-1}$ insulin (Life Technologies) at 37 °C with 4% CO_2. Cell-line identity was confirmed by short-tandem-repeat analysis and the cells were regularly tested for mycoplasma (Venor GeM mycoplasma detection kit, Sigma).

PE1 was a primary cell line from the pleural effusion of a patient with triple-negative (ER$^-$PR$^-$HER2$^-$) breast cancer. The pleural effusion was collected under the Dana–Farber Harvard Cancer Center (DF/HCC) Institutional Review Board (IRB) protocol no. 93-085 following written informed consent and used in the lab in compliance with DF/HCC IRB protocol no. 14-400 approved for the use of de-identified tissue samples. The study is compliant with all of the relevant ethical regulations regarding research involving human participants.

Xenograft experiments. All of the animal procedures were approved by DFCI IACUC and performed according to the DFCI protocol no. 11-023. The study is compliant with all of the relevant ethical regulations regarding animal research. Female NOG (NOD.Cg-$Prkdc^{scid}\,Il2rg^{tm1Sug}$/JicTac) mice (Taconic) were used at 4–6 weeks of age. For the tumour growth assays, a total volume of 50 μl containing 3×10^6 cells in 50% Matrigel (BD Biosciences) in DMEM media (Gibco) was injected into the mammary fat pad. Tumour growth was measured weekly with a caliper. In some experiments, metastatic growth was followed by weekly bioluminescence imaging measurements of luciferase activity using an IVIS imaging system. The size of the macrometastasis was determined by Living Image software (PerkinElmer) analysis of the areas of bioluminescence in the images. Regions of interest were automatically selected by the software and the area of each individual region of interest was measured. The parental tumour region of interest was excluded from the analysis. Blood samples at early time points were collected via retro-orbital bleeding. For the in vivo inducible expression of IL11 and FIGF, the mice were switched to a DOX diet 24 h after a mammary-fat-pad injection.

FACS analysis. For all experiments, blood was freshly collected and processed immediately, whereas bone marrow was viably frozen after erythrolysis. Tumour and lung samples were digested to single-cell suspensions with collagenase type IV (Worthington, cat no. LS004189)/DNase I digestion buffer ($2\,mg\,ml^{-1}$ and $0.02\,mg\,ml^{-1}$, respectively, in DMEM with 10% FBS and 1% Pen-Strep). The digestion was performed at 37 °C with stirring for 1 h for tumour tissues and 0.5 h for lung tissues. The digestion was terminated with DMEM containing 10% FBS, and the cells were washed and stained immediately. The red blood cells were lysed with lysis buffer (BD Pharm Lyse, BD Biosciences). After washing with FACS buffer (2% FBS in PBS and 2 mM EDTA), cells were stained with antibody cocktail 1 or 2 (Supplementary Table 2) for 20–30 min on ice and washed twice. The final cell pellet was resuspended in FACS buffer with or without 7AAD for viability detection. The analysis was performed on a BD FACS Canto system (BD Biosciences). Gating and analysis of the peripheral blood mononuclear cell subpopulations were performed using FlowJo software.

In vivo neutrophil depletion. The mice were treated with 12.5 μg anti-Ly6G (BioCell) or isotype-control antibody (BioCell) in a 100 μl volume via intra-peritoneal injections 24 h post mammary-fat-pad injection. The treatment was administered daily for 28 d. The on-target effect was measured after two weeks of treatment by FACS analysis of the peripheral blood mononuclear cells in the mouse blood.

Immunofluorescence and immunohistochemistry. After collection, xenograft samples and mouse organs were fixed in buffered formalin overnight and stored in 70% ethanol before being processed into formalin-fixed paraffin-embedded blocks by the DFCI Pathology Core. Staining was performed on 5-μm sections. Before staining, the slides were baked at 65 °C and a standard deparaffinization protocol using xylene and ethanol washes was used. If required, antigen retrieval was performed using citrate buffer (pH 6; Invitrogen) or Target Retrieval solution (pH 9; Dako) for 20 min in a steamer. Some antibodies performed better after a 10-min permeabilization step using 0.5% Triton X-100 (Sigma Aldrich) in PBS. The specific antibodies and dilutions used are listed in Supplementary Table 2. Incubation with primary antibodies diluted with PBST buffer (0.05% Tween 20 in PBS) in 5% goat serum was performed overnight at 4 °C. After three washes with PBST, the secondary antibody was applied to the slides and incubated for 1 h at room temperature. For immunohistochemical analysis, the ABC peroxidase system (Vectastain, ABC System Vector Laboratories) was used with 3,3′-diaminobenzidine as the colorimetric substrate. Counterstaining was performed using Harris haematoxylin and Scott's tap water substitute (Leica). Automated quantification of individual positive cells was performed using ImageJ macros (available on request) by scoring the total cell numbers based on nuclear staining and strongly positive cells across five different areas of a given specimen.

Following immunofluorescent staining, the slides were counterstained with DAPI (Vectashield Hard Set Mounting Medium with DAPI) and stored for at least 24 h at −20 °C to reduce autofluorescence. Images from multiple areas of each sample were acquired using a Nikon Ti microscope attached to a Yokogawa spinning-disk confocal unit using a 603 plain apo objective and an OrcaER camera controlled by Andor iQ software or with a Leica SP5 confocal scanning microscope.

Stroma and epithelial cell fractionation and RNA-seq. *Cell fractionation.* To separate subpopulations from polyclonal tumours, the following MDA-MB-468 cell line derivatives with fluorescent tags were used: neutral clones (expressing CFP and Thy1.1), and IL11–mCherry and FIGF–GFP clones. Polyclonal tumours were dissociated to a single-cell suspension by digestion in collagenase/hyaluronidase digestion buffer ($2\,mg\,ml^{-1}$ each in DMEM with 10% FBS and 1% Pen-Strep). After washing with PBS, the samples were resuspended in FACS buffer (2% FBS in PBS) and the populations expressing CFP, mCherry, GFP and Thy1.1 were sorted into FACS buffer, pelleted and stored at −80 °C. The negative cell population was also collected, as this constitutes the tumour stroma. The same procedure was performed on metastatic lungs, except all labelled cells were collected as a group due to their overall small numbers in metastatic lungs.

mRNA extraction and library preparation. Messenger RNA selection and whole transcriptome amplification were performed as previously described with some modifications[39,40]. A primer 5′-phosphorylated oligo-dT24 (pdT24) was used to select mRNA from total RNA. For denaturation and primer annealing, total RNA was mixed with dNTPs (25 mM) and pdT24, incubated in a PCR machine preheated to 68 °C for 5 min and then immediately placed on ice. Reverse transcription was performed with the addition of the First-Strand buffer, dithiothreitol, RNase inhibitor and $10\,U\,\mu l^{-1}$ Superscript reverse transcriptase III. Double-stranded cDNA was generated by the addition of 30 μl of a mix containing 10×Second Strand buffer, dNTPs (25 mM), *Escherichia coli* DNA ligase ($10\,U\,\mu l^{-1}$), RNaseH and DNA polymerase ($10\,U\,\mu l^{-1}$), followed by a 3 h incubation at 16 °C and 5 min at 70 °C. The product was purified with the Genomic DNA Clean and Concentrator kit (Zymo Research). Next, DNA blunt-ending, 5′-end phosphorylation and ligation were performed with an End-It DNA end-repair kit (Epicentre) and T4 DNA ligase (Epicentre). The product was directly amplified by adding a mixture of Reaction Buffer and DNA polymerase from the REPLI-g UltraFast kit (Qiagen) at a 15:1 ratio. The amplification reaction was carried out at 30 °C for 2 h. The amplified product was purified using the same Genomic DNA Clean and Concentrator kit (Zymo Research). Typically 5–10 μg of mRNA-derived cDNA amplicons were obtained when eluted in 50 μl elution buffer. The amplicons were then evaluated for quantity, coverage and purity by PCR using a set of primers for housekeeping genes (*GAPDH*: forward, 5′-GTCTCCTCTGACTTCAACAGCG-3′ and reverse, 5′-ACCACCCTGTTGCTGTAGCCAA-3′; *B2M*: forward, 5′-CCACTGAAAAAGATGAGTATGCCT-3′ and reverse, 5′-CCAATCCAAATGCGGCATCTTCA-3′), cDNA 5′ ends (*CERB5*: forward, 5′-GTCAGTGAACTCCAGCATCATGG-3′ and reverse, 5′-GTGGTGAGTCAATGCAGCCTTC-3′) and genomic DNA (*10p*: forward, 5′-GTTCTGCTGCCTCTACACAGG-3′ and reverse, 5′-ATCCTTCTGTGAACTCTCAAATTC-3′). A Nextera XT DNA sample preparation kit (Illumina) was used according to the manufacturer's recommendations with minor modifications to construct cDNA libraries. Briefly, cDNA amplicons were fragmented and barcoded by tagmentation using Tn5 DNA transposase. The tagmented cDNA was amplified with a limited-cycle PCR program. During the PCR step, indices 1 and 2 were added to enable cluster generation and multiplexed sequencing. The extension time was increased from 30 s to 60 s. The amplified cDNA libraries were purified and cleaned using 0.6×Agencourt AMPure XP (Beckman Coulter) and fresh ethanol, and eluted in TE buffer. The quality of the sequencing libraries was assessed using a high-sensitivity DNA chip (Agilent). Finally, the libraries were deep sequenced using a HiSeq2000 sequencer (Illumina).

Clustering based on differentially expressed genes. We analysed the stroma from primary tumours and metastatic lungs separately. First, we normalized for sequencing depth and applied the variance stabilizing transformation, both implemented in DESeq2. The top 1,000 most variable genes were used to cluster the samples based on the expression of these genes (Euclidean distance and complete clustering). The heat map itself depicts relative variance-stabilizing-transformed values across samples (the difference between the value of the sample for a given gene and the mean expression of the given gene across all samples) and hence negative values were obtained. Note that this way of displaying the expression does not affect the sample clustering when compared with clustering based on expression-correlation values.

RNA-seq of lungs. Lungs from tumour-bearing mice with and without DOX were isolated and dissociated into single-cell suspensions as described above. The lungs from five mice were pooled for each group with two replicates. Staining for IL11R and PLXDC2 was performed as described in the 'FACS analysis' section. IL11R$^-$, IL11R$^+$PLXDC2$^-$ and IL11R$^+$PLXDC2$^+$ populations were sorted by FACS into RLT

buffer with β-mercaptoethanol (Qiagen). Total RNA was then isolated as described by the manufacturer using the RNeasy mini kit (Qiagen). The extraction of mRNA and library preparation were performed as described above.

RNA-seq analysis. Raw sequencing data were de-multiplexed into fastq files with bcl2fastq software (Illumina) based on unique base pairs. Quality checks of the called bases of the reads were performed using the FastQC tool (Babraham Bioinformatics). Tumour-sample fastq files were trimmed to 40 bases. The datasets of tumour origin were aligned to the human reference GRCh37/hg19 genome and datasets of stroma origin were aligned to the mouse reference GRCm38/mm10 genome using the STAR RNA-seq aligner (v.STAR_2.5.1b) for both. Two-pass mapping was performed using the following modified parameters: —outSAMstrandField intronMotif, —outFilterMultimapNmax 20, —alignSJoverhangMin 8, —alignSJDBoverhangMin 1, —outFilterMismatchNmax 999, —outFilterMismatchNoverLmax 0.1, —alignIntronMin 20, —alignIntronMax 1,000,000, —alignMatesGapMax 1,000,000, —outFilterType BySJout, —outFilterScoreMinOverLread 0.33, —outFilterMatchNminOverLread 0.33, —limitSjdbInsertNsj 1,200,000, —chimSegmentMin 15, —chimJunctionOverhangMin 15, —twopassMode Basic. The samples were filtered to remove contamination (mouse contamination in tumour samples and human contamination in stroma samples). Reads that uniquely mapped to the target genome were kept along with uniquely mapped reads that had significantly better alignment scores in the target genome compared with the contamination genome. Read counts for individual genes were generated using the htseq-count script of the HTSeq framework (v.0.6.1p1) using modified parameters (—stranded no) and the hg19/mm10 refGene annotation file available at the UCSC Genome. For the stroma samples, reads for technical duplicates were added together. The DESeq2 R package v.1.16.1 was used to generate differential expression gene lists, with a fold-change ≥ 2 and adjusted $P \leq 0.05$.

scRNA-seq. Blood, tumours and lungs were isolated from the test groups and tumour-naive animals. The tissues were dissociated into single-cell suspensions as described above. Staining for CD45 was performed as for the FACS analysis of the peripheral blood mononuclear cells. Samples were pooled across three animals per group and CD45$^+$ populations were sorted by FACS into PBS with 2% RNase-free BSA (Ambion). The cells were then processed according to the 10xGenomics sample preparation protocol (Chromium Single Cell 3′ v2 Reagent Kit). Two thousand cells were targeted for each sample.

Analysis of scRNA-seq data. *Generating counts.* Cell Ranger v.1.2 was used to pre-process the Chromium single cell 3′ RNA-seq output. Specifically, 'cellranger mkfastq' was used to demultiplex the samples and produce FASTQ files, 'cellranger count' was used to align and filter the FASTQ files individually, generating gene-barcode matrices, and 'cellranger aggr' was used to aggregate all of the samples into a single combined matrix, normalizing the sequencing depth across runs. The resulting data were analysed with the R (v.3.4.) package Seurat v.2.1.0 (ref. [41]).

Quality control and normalization. The data were filtered by selecting all genes that were expressed in at least a single cell and all cells with at least 100 and at most 5,000 expressed genes, and with a maximum of 0.1% of mitochondrial DNA. This amounted to keeping a total of 18,203 genes and 7,704 cells. Gene-expression measurements for each cell were normalized to the total expression, multiplied by a factor of 10,000 and the result was log transformed (normalization method LogNormalize in Seurat).

Variable genes across cells. Genes that were highly variable across cells were identified with the Seurat function 'FindVariableGenes' with the following parameters: mean.function = ExpMean, dispersion.function = LogVMR, x.low.cutoff = 0.0125, x.high.cutoff = 3 and y.cutoff = 0.5. This amounted to selecting 2,571 genes for downstream dimensionality reduction and clustering. The number of UMIs (unique molecular identifiers) and the percentage of mitochondrial DNA across all cells were regressed against each gene, and the resulting residuals were further scaled and centred.

Dimensionality reduction. Principal component analysis was performed on the scaled data. After regressing out the unwanted sources of variation, the PCA component explaining the highest percentage of variation did not correlate with the number of expressed genes (Pearson's correlation of 0.01). Cells were clustered based on the PCA scores of the first 20 principal components, selected as strongly significant after running the resampling test implemented in the Seurat function 'JackStraw'. The clustering of the data was based on the graph-based procedure implemented in the Seurat function 'FindClusters', for which a resolution parameter of 0.3 was chosen. Briefly, the underlying clustering methodology includes constructing a K-nearest neighbour on the Euclidean distances among cells in the PCA space and refining edge weights based on the Jaccard distance between any two cells. To visualize the data, which clustered into 14 clusters, tSNE was used. One cluster consisting of 712 cells was identified to represent tumour cells by its high expression of *EGFR* and numerous keratins, including *KRT14* and *KRT18* (Supplementary Fig. 6c). These 712 cells were removed from the dataset.

The new data was clustered again using a resolution parameter of 0.2, which resulted in 12 clusters.

Differential expression. Genes that were differentially expressed between the clusters were identified using the Wilcoxon rank-sum test, as implemented in the Seurat function 'FindAllMarkers', with a minimum fraction of expression across cells of 0 (min.pct) and an expression threshold on the differences in log fold changes of 0.25 (logfc.threshold).

Identification of cell type. Cell clusters were assigned to cell types based on the expression of classical cell-type-specific markers and the similarity between the top 30 cluster-specific genes and the ImmGen consortium datasets (http://rstats.immgen.org/MyGeneSet/)[42].

Identification of secreted factors. We used the lists of the top 1,000 markers that were differentially expressed between the MSC cluster and rest of the cells from +DOX mice or −DOX mice. The NABA_SECRETED_FACTORS gene set from the GSEA database (http://software.broadinstitute.org/gsea/msigdb/index.jsp) was used to select for genes encoding secreted factors in those lists.

Mathematical modelling of metastasis onset. To estimate the timing of the first metastatic cell emigrating from the primary tumour, we created a birth–death process model of cell growth in which an individual cell lives for an exponentially distributed amount of time before splitting into two daughters[43,44]. In this model, the initiating population seeded within the primary site consists entirely of so-called 'type-0' cells. We considered situations in which type-0 cells have already evolved the potential to disseminate to other organs, such as lymph nodes, lung and bone, based on the fact that no metastatic driver genes have been reported. Type-0 cells divide at the rate r_0 and die at the rate d_0 per week, and may leave the primary site to found a new metastasis at a distant site at the rate q per week. Each export event results in a new metastatic colony. Metastasized cells are referred to as 'type-1' cells and may go extinct or grow with a division and death rate of r_1 and d_1, respectively, per week. The total number of metastatic sites until time t is denoted by $I(t)$. During each elementary time step of the stochastic process, cells may divide, die or leave the primary site to initiate a colony elsewhere.

Numerical simulations. We performed numerical simulations of the stochastic process. Here, w and x_i denote the number of type-0 cells and type-1 cells of the ith metastatic site, respectively. A change in w and x_i can occur by cell division, cell death or departure of a cell from a primary site. Initially, the population of cancer cells is initialized with 3×10^6 type-0 cells and no type-1 cells, corresponding to 3×10^6 cells injected into mice for the in vivo experiments. Our model considers all possible events in the stochastic simulation—production and death of a type-0 or type-1 cell, and metastasis of a type-0 cell to a new site. The probability of each event occurring is its rate of occurrence normalized to the sum of the rates of all possible events, given by $\Gamma = (r_0 + d_0 + q)w + \Sigma_i(r_1 + d_1)x_i$. The timing of one event is exponentially distributed with mean $1/\Gamma$. The transition probabilities between states in the stochastic simulation are determined as follows: the probabilities that the number of each type cell increase by one are

$$\Pr\left[(w, x_1, x_2, \ldots, x_I) \rightarrow (w+1, x_1, x_2, \ldots, x_I)\right] = r_0 w / \Gamma$$
$$\Pr\left[(w, x_1, x_2, \ldots, x_i, \ldots, x_I) \rightarrow (w, x_1, x_2, \ldots x_i + 1, \ldots, x_I)\right] = r_1 x_i / \Gamma \qquad (1)$$

Export of a type-0 cell to a new metastatic site increases the number of metastatic cells by one at the new site $I + 1$ and decreases the number of type-0 cells by one; the probability of this event occurring is

$$\Pr\left[(w, x_1, x_2, \ldots, x_I) \rightarrow (w-1, x_1, x_2, \ldots, x_I, 1)\right] = wq / \Gamma \qquad (2)$$

The probabilities that the numbers of type-0, -1 and -2 cells in the ith metastatic site decrease by one are

$$\Pr\left[(w, x_1, x_2, \ldots, x_I) \rightarrow (w-1, x_1, x_2, \ldots, x_I)\right] = d_0 w / \Gamma$$
$$\Pr\left[(w, x_1, x_2, \ldots, x_i, \ldots, x_I) \rightarrow (w, x_1, x_2, \ldots, x_i - 1, \ldots, x_I)\right] = d_1 x_i / \Gamma \qquad (3)$$

Based on the framework described in equations (1)–(3), we performed independent runs of the stochastic process, either until all cells go extinct, $w = x = \Sigma_i y_i = 0$, or for a total time of six weeks after the cancer cells were seeded. Our main goal was to determine the distribution of the size of the first metastasis at week six and the size of primary site where the first metastasis was detected.

Parameter estimation. We explored the use of mixed-effects models for the datasets where repeated tumour-volume measurements were made over time in each mouse. For both the primary and metastatic sites in each group of IL11 and polyclonal tumour-bearing mice, the growth rates were determined by analysing the number of cancer cells using a log-transformed linear regression. Given that 3×10^6 cells were injected into the mice at the beginning of the experiment, we set the number of cancer cells in the primary site at week zero to 3×10^6. The estimated slope of the regression then provides the estimated growth rate. Tumour sizes

utilizing two measurement methods were available for the primary site: tumour diameter by caliper measurements and the largest cross-sectional area of a tumour by bioluminescence detection. For the metastatic sites, caliper measurements were not available because the tumour sizes were too small to be measured; therefore, only cross-sectional-area measurements were performed by bioluminescence detection. Based on the idea that growth estimation based on caliper measurements is more accurate than bioluminescence detection methods due to oversaturation of large lesions, we calculated the growth rates of metastatic sites by converting bioluminescence to caliper measurements using primary-site data. Specifically, we calculated the ratio of the estimated growth rates of the primary sites by the two aforementioned measurement methods, caliper measurement and bioluminescence detection, and multiplied it by the growth rates of the metastatic sites estimated by bioluminescence. We converted the tumour volume to cell numbers with the assumption of a spherical shape (10^9 cells occupy a volume of $1\,cm^3$). We performed 100 simulations per metastatic rate q (q in $[10^{-10}, 10^{-5}]$) to estimate the parameters of our model.

MBCP cohort analysis. All patients for whom we analysed tumour tissue provided written informed consent for research biopsies and transcriptome sequencing of tumour RNA, as approved by the DF/HCC IRB (protocol no. 05-246), before any study procedures were performed. The study is compliant with all of the relevant ethical regulations regarding research involving human participants. The cohort consists of 51 primary tumours (formalin-fixed paraffin-embedded samples) and matching 155 metastatic lesions (mostly frozen tissue) enriched in ER+ subtype. Because of the difference in tissue-preservation methods between primary and metastatic lesions, batch effects cannot be ruled out when comparing primary and metastatic lesions[45]. Total RNA was extracted using a RNeasy mini kit (Qiagen) and libraries for RNA-seq were prepared using the TrueSeq RNA sample preparation kit protocol adapted for the Sciclone liquid handler (PerkinElmer). Sequencing was performed on an Illumina Nextseq500. Reads were aligned to the hg19 human genome using STAR v.2.5.1 (ref. [46]) followed by Transcript assembly using cufflinks v.2.2.1 (ref. [47]). Quality-control steps were done using STAR v.2.5.1 and RseQC v.2.6.2 (ref. [48]). We assessed each sample on metrics of mappable reads, percentage of rRNA reads, gene body coverage, and junction saturation and insert_size for paired ends to determine which samples were of adequate quality. Signatures of IL11/FIGF-activated neutrophils and MSCs from blood and lungs were derived from differentially expressed gene lists from our scRNA-seq analysis. We aimed to construct the signatures using the top 150 differentially expressed genes, unless fewer genes presented with an adjusted $P < 0.05$. The signature of the cancer cells was derived from the top 150 genes that were differentially expressed between the Thy1.1 neutral clone sorted from polyclonal tumour and the parental cell line tumour. All signatures are presented in Supplementary Table 6. Signature analysis on the MBCP and TCGA data was performed using GSVA[49].

Statistics and reproducibility. The sample sizes for the mouse experiments were determined based on previous studies and pilot experiments to ensure that sufficient power could be obtained[7]. Unless otherwise stated, statistical analyses were performed using GraphPad Prism software or R. Box-and-whisker plots show the mean (midline), and 25th–75th (box) and 5th–95th percentiles (whiskers). Most experiments in this study were performed independently at least twice and

all attempts at replication of the results presented in this study were successful. Biological replicates were included in all experiments.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

RNA-seq and scRNA-seq data have been deposited in the NCBI GEO database with the accession number GSE109281. The publicly available subset of the MBCP RNA-seq data can be found in GEO database with the accession number GSE121411. Primary breast cancer data, where indicated, were derived from the TCGA dataset (http://cancergenome.nih.gov/). The source data for Figs. 1,3,4,6 and Supplementary Fig. 1–6 have been provided as Supplementary Table 1. All data supporting the findings in this study are available from the corresponding author on request.

Code availability

The mathematical modelling code is available from the corresponding author on request.

References

39. Pan, X. et al. Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proc. Natl Acad. Sci. USA* **110**, 594–599 (2013).
40. Guo, S. et al. Nonstochastic reprogramming from a privileged somatic cell state. *Cell* **156**, 649–662 (2014).
41. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
42. Heng, T. S. et al. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
43. McDonald, T. O. & Michor, F. SIApopr: a computational method to simulate evolutionary branching trees for analysis of tumor clonal evolution. *Bioinformatics* **33**, 2221–2223 (2017).
44. Haeno, H. et al. Computational modeling of pancreatic cancer reveals kinetics of metastasis suggesting optimum treatment strategies. *Cell* **148**, 362–375 (2012).
45. Chen, R. et al. Robust transcriptional tumor signatures applicable to both formalin-fixed paraffin-embedded and fresh-frozen samples. *Oncotarget* **8**, 6652–6662 (2017).
46. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
47. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
48. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
49. Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7 (2013).

nature research

Corresponding author(s): Kornelia Polyak

Last updated by author(s): Apr 24, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	Statistical analyses were performed using GraphPad Prism 7 software or R (v.3.4.). RNAseq experiments were analyzed using bcl2fastq software (Illumina), FastQC tool (Babraham Bioinformatics), STAR RNA-Seq aligner (version STAR_2.5.1b), HTSeq framework (version 0.6.1p1), DeSeq2 R package (v1.16.1). Single cell RNAseq data was analyzed using the Cell Ranger software (v.1.2.) and Seurat R package (v.2.1.0). Human cohort analysis was performed using STAR (v2.5.1), cufflinks v2.2.1, RseQC v2.6.2. GSVA and GSEA were also used. Quantification of bioluminescence was performed using Living Image (perkin Elmer) software Quantification of immunohistochemical and fluorescent staining was performed using ImageJ (v.2.0.0-re-69/1.52n). FACS results were analyzed using FlowJo software (v.10)..

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-seq and scRNA-seq data have been deposited in the NCBI GEO database with the accession number GSE109281. These data are associated with Figures 2, 5 and 6 and Supplemental Figures 2 and 6. Publicly available subset of the MBC Project RNA-seq data is available in GEO database with the accession number GSE121411.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The number of animals per each group was determined based on our prior study using the same model (Marusyk et al., Nature 2014). We estimated that 5 animals per group would account for variability in tumor growth.
Data exclusions	In our single cell RNAseq gene expression analysis we excluded cells that expressed less than 100 genes, and cells with ,pre than 0.1% of mitochondrial DNA. This allows to eliminate dead cells and cell fragments from final analysis.
Replication	All attempts at replication were successful. All animal and RNAseq experiments contained biological replicates.
Randomization	In neutrophil depletion experiment, animals injected with tumor cells were randomized to different treatment groups.
Blinding	Imaging of metastatic lesions by bioluminescence and with IHC staining was performed by technical assistants, who were blinded to the treatment groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Cytokeratin LMW mouse IgM DAKO 35βH11 1:100 Cytokeratin HMW mouse IgG1 DAKO 34βE12 1:100 GFP rabbit Cell Signaling 2956S 1:100 V5-tag mouse IgG2a Invitrogen R960-25 1:200 CD31 rabbit Neomarkers RB-10333-P 1:50 LYVE1 rabbit Angio-Proteomie 103-PA50AG 1:500 HA-tag rabbit mAb Cell Signaling 3724 1:400 IL11R (4D12) mouse IgG Santa Cruz Biotech. Sc-130920 1:100 CD45 mouse IgG BD Biosciences BD 550539 1:50 TEM8 rabbit Abcam ab21270 1:50 PLXDC2 rabbit Abcam ab67225 1:50 Anti-rabbit-biotin goat ThermoFisher 65-6140 1:100 Anti-mouse-biotin goat Vector Laboratories BA-9200 1:100 Anti-mouse-AlexaFluor488 goat ThermoFisher R37120 1:100 Anti-rabbit-AlexaFluor555 goat ThermoFisher A27039 1:100 Anti-rabbit-AlexaFluor488 goat ThermoFisher A11034 1:100 PB CD45 rat IgG2b BioLegend BL 103126 1:200 Biotin Thy1.1 Mouse IgG1 BioLegend BL 202510 1:500 FITC CD11b rat IgG2b BioLegend BL 101205 1:200 PE Ly6G rat IgG2a BioLegend BL 127617 1:300 APC Ly6C rat IgM BD Biosciences BD 560595 1:300 APC-Cy7 Siglec F rat IgG2a BD Biosciences BD 565527 1:200

PE-Cy7 F4/80 rat IgG2a BioLegend BL 123113 1:200
PE CD11c armenian hamster IgG BioLegend BL 117307 1:300
APC-Cy7 Sca1 rat IgG2a BioLegend BL 108126 1:200
APC FceRI armenian hamster IgG BioLegend BL 134315 1:100
AF488 cKit rat IgG2b BioLegend BL 105815 1:100
PE-Cy7 B220 rat IgG2a BioLegend BL 103221 1:250
APC B220 rat IgG2a BioLegend BL 103212 1:300
FITC CD11b rat IgG2b BioLegend BL 101205 1:200
PE Gr1 rat IgG2b BioLegend BL 108407 1:100
FITC IL11Ralpha mouse IgG SantaCruz Biotechnology SC 130920 1:2
Streptavidin Alexa Fluor 647 ThermoFisher S21374 1:100
Zenon-Alexa Fluor-anti-mouse 555   goat ThermoFisher
 Z25005 5 L/ g
Zenon-Alexa Fluor-anti-rabbit 488 goat ThermoFisher
 Z25302 5 L/ g
Cytokeratin LMW mouse IgM DAKO 35βH11 1:100
Cytokeratin HMW mouse IgG1 DAKO 34βE12 1:100
GFP rabbit Cell Signaling 2956S 1:100
V5-tag mouse IgG2a Invitrogen R960-25 1:200
CD31 rabbit Neomarkers RB-10333-P 1:50
LYVE1 rabbit Angio-Proteomie 103-PA50AG 1:500
HA-tag rabbit mAb Cell Signaling 3724 1:400
IL11R   (4D12) mouse IgG Santa Cruz Biotech. Sc-130920 1:100
CD45 mouse IgG BD Biosciences BD 550539 1:50
TEM8 rabbit Abcam ab21270 1:50
PLXDC2 rabbit Abcam ab67225 1:50
Anti-rabbit-biotin goat ThermoFisher 65-6140 1:100
Anti-mouse-biotin goat Vector Laboratories BA-9200 1:100
Anti-mouse-AlexaFluor488 goat ThermoFisher R37120 1:100
Anti-rabbit-AlexaFluor555 goat ThermoFisher A27039 1:100
Anti-rabbit-AlexaFluor488 goat ThermoFisher A11034 1:100
PB CD45 rat IgG2b BioLegend BL 103126 1:200
Biotin Thy1.1 Mouse IgG1 BioLegend BL 202510 1:500
FITC CD11b rat IgG2b BioLegend BL 101205 1:200
PE Ly6G rat IgG2a BioLegend BL 127617 1:300
APC Ly6C rat IgM BD Biosciences BD 560595 1:300
APC-Cy7 Siglec F rat IgG2a BD Biosciences BD 565527 1:200
PE-Cy7 F4/80 rat IgG2a BioLegend BL 123113 1:200
PE CD11c armenian hamster IgG BioLegend BL 117307 1:300
APC-Cy7 Sca1 rat IgG2a BioLegend BL 108126 1:200
APC FceRI armenian hamster IgG BioLegend BL 134315 1:100
AF488 cKit rat IgG2b BioLegend BL 105815 1:100
PE-Cy7 B220 rat IgG2a BioLegend BL 103221 1:250
APC B220 rat IgG2a BioLegend BL 103212 1:300
FITC CD11b rat IgG2b BioLegend BL 101205 1:200
PE Gr1 rat IgG2b BioLegend BL 108407 1:100
FITC IL11Ralpha mouse IgG SantaCruz Biotechnology SC 130920 1:2
Streptavidin Alexa Fluor 647 ThermoFisher S21374 1:100
Zenon-Alexa Fluor-anti-mouse 555   goat ThermoFisher
 Z25005 5 L/ g
Zenon-Alexa Fluor-anti-rabbit 488 goat ThermoFisher
 Z25302 5 L/ g

Validation	Antibodies used for FACS staining were validated on NOG mouse tissues, and the obtained percentages corresponded well to the expected values. IHC and IF antibodies were validated on appropriate FFPE mouse tissue and xenografts.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	MDA-MB-468 was from ATCC. PE1 is a primary line from pleural effusion of a metastatic breast cancer patient.
Authentication	MDA-MB-468 cell line was tested by STR and SNP array profiling by the DFCI Molecular Biology core facility.
Mycoplasma contamination	MDA-MB-468 cell line were routinely tested for mycoplasma and murine pathogen contamination. No contamination was found at any time point during the course of the study.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	4–6-weeks old female mice NOG (NOD.Cg-Prkdcscid Il2rgtm1Sug/JicTac) mice (Taconic) were used
Wild animals	Study did not involve wild animals.
Field-collected samples	Study did not involve field-collected samples.
Ethics oversight	All animal procedures were approved by DFCI ACUC and performed according to DFCI protocol #11-023. All patients participating in the MBC Project provided a written informed consent for research metastatic biopsies and genomic profiling of tumor RNA. The study was approved by the Dana-Farber/Harvard Cancer Center Institutional Review Board (DF/HCC Protocol 05-246).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	FACS analysis was performed on blood, primary tumors and metastatic lungs. Blood samples were collected by retroorbital bleeding or cardiac puncture, subject to erythrolysis and stained according to standard protocols. Lung and tumor samples were digested into single cell suspension with collagenase/hyaluronidase/DNase mix, washed and stained accordingly.
Instrument	FACS analysis was performed on BD FACS Canto II. Cell sorting for bulk and single-cell RNAseq experiments was performed on BD FACS Aria II.
Software	FlowJo software (v10) was used to analyze all cytometry data.
Cell population abundance	CD45 cell sorting was previously validated in the lab by RNAseq of known positive and negative cell populations. IL11RA and PLXDC2 cell sorting were validated by flow cytometic analysis and correlation to RNAseq data from cell lines with and without expression of these proteins.
Gating strategy	FSC/SSC gates were chosen to eliminate debris and doublets. For cell sorting experiment for single-cell RNAseq , dead cells were eliminated by excluding 7AAD+ cells. For cell sorting experiment for bulk RNAseq , live cells were gated based on FSC and SSC. Within a single experiment, gating was chosen based on random sample gating, before a single gate was applied to all plots for final analysis. An example of gating strategy is presented in the Supplementary Fig. 1, 3, and 6.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.